

## 基于云计算的内容感知的社交网络可视化

社交媒体包含大量用户而且用户间的联系构成了庞大而复杂的网络。从用户的角度来说，最希望看到的是图形化的整个社交网络图，既能看到全局结构也能深入详细地查看局部细节。但是这是明显的大数据问题，此类数据的可视化仍然缺少有效的方法。

云计算是目前人类处理大规模数据最有效的方法，不仅硬件性价比高，而且 MapReduce 的计算方式直观而且有效。然而云计算的研究仍处于基础研究阶段，云计算与社交网络可视化相结合的研究查无文献。

图绘制是社交网络可视化的基础，而图绘制的基础又是图布局。以力引导模型为代表的图绘制方法逐渐成为图绘制的主要方法。力引导模型不依赖于应用，以力学、几何学基本原理为基础，依据力引导模型建立的算法直观易懂而且可视化效果令人满意。基于多级抽象的多尺度技术已经将图绘制的节点提高到上亿规模，然而即使采用 GPU、高性能集群等并行技术绘制性能仍然有待提高（AT&T 1 亿节点需要 88.5 小时）。

社交网络具有一些特殊的性质，如 Degree Centrality, Betweenness Centrality, Closeness, Authority 等，根据这些性质能轻松对整个社交网络进行划分。根据 MapReduce 的基本原则，只要能够合理地划分节点，就可以在不同的机器上完成分析、可视计算等工作。在每台机器上完成局部视图的绘制之后，最终合并成全局结构图，为用户呈现多尺度社交网络图。

解决问题的关键在于两个方面：一是如何划分节点，让不同的机器任务比较均衡；二是要尽量降低网络通信开销，减少数据复制。内容感知(Content Aware)就是让云计算平台知道需要处理的数据内部信息，依据数据内容合理划分数据，而不是按照数据节点的大小(64M)均匀拆分。根据内容划分，比如按照社区划分，可以让同一台机器上的数据通信多，而不同机器间的通信尽量少。

以上是最近总结的一些基本思路，但是由于对云计算相关知识不熟悉，最近正在恶补，同时参加了张老师的一个讨论班一起学习。

### 力引导模型发展历史

本周还系统地学习了力引导图的绘制算法。最早的力引导模型是 Tutte 在 1963 提出的中心点方法，将节点之间的边当作是弹簧，并假设弹簧的理想长度为 0，由此计算顶点之间的排斥力，产生最终布局。1984 年 Eades 改进了力引导模型，不仅考虑排斥力而且考虑吸引力。将顶点想象成铁环，连接的边带上弹簧构成力学系统，铁环间受弹簧拉力相互靠近，同时为了避免过于靠近又引入电磁力。通过计算顶点间的合力确定最终的布局。1991 年 Fruchterman 和 Reingold 进一步完善了该模型。1989 年 Kamada 和 Kawai 提出了另一种计算思路，理论距离方法，假设任意两个顶点间存在一个理论最优的几何距离，通过将初始的位置向理论距离靠近进行优化求解，最终形成比较好的图布局。然而这些算法的复杂度较高，无法应用于顶点规模较大的情形。Hadany 和 Harel 在 2001 年做出了突破，引入多尺度技术结合力引导模型成功绘制大于 1000 个顶点的图。所谓多尺度技术其实是对图的多级抽象，在粗粒度级别只考虑与图结构有关的特征，而在细粒度级别才去关注局部细节特征。在此基础上，更多顶点和更快速的图绘制算法相应提出，最新的大图绘制算法应该是胡一凡提出的 SFD 算法。